

## RESEARCH ARTICLE

## Open Access

# Positive selection at high temperature reduces gene transcription in the bacteriophage $\phi$ X174

Celeste J Brown<sup>1\*</sup>, Luyi Zhao<sup>1</sup>, Kelsie J Evans<sup>1</sup>, Dilara Ally<sup>1,2</sup>, Amber D Stancik<sup>1</sup>

## Abstract

**Background:** Gene regulation plays a central role in the adaptation of organisms to their environments. There are many molecular components to gene regulation, and it is often difficult to determine both the genetic basis of adaptation and the evolutionary forces that influence regulation. In multiple evolution experiments with the bacteriophage  $\phi$ X174, adaptive substitutions in *cis*-acting regulatory sequences sweep through the phage population as the result of strong positive selection at high temperatures that are non-permissive for laboratory-adapted phage. For one *cis*-regulatory region, we investigate the individual effects of four adaptive substitutions on transcript levels and fitness for phage growing on three hosts at two temperatures.

**Results:** The effect of the four individual substitutions on transcript levels is to down-regulate gene expression, regardless of temperature or host. To ascertain the conditions under which these substitutions are adaptive, fitness was measured by a variety of methods for several bacterial hosts growing at two temperatures, the control temperature of 37°C and the selective temperature of 42°C. Time to lysis and doublings per hour indicate that the four substitutions individually improve fitness over the ancestral strain at high temperature independent of the bacterial host in which the fitness was measured. Competition assays between the ancestral strain and either of two mutant strains indicate that both mutants out-compete the ancestor at high temperature, but the relative frequencies of each phage remain the same at the control temperature.

**Conclusions:** Our results strongly suggest that gene transcription plays an important role in influencing fitness in the bacteriophage  $\phi$ X174, and different point mutations in a single *cis*-regulatory region provided the genetic basis for this role in adaptation to high temperature. We speculate that the adaptive nature of these substitutions is due to the physiology of the host at high temperature or the need to maintain particular ratios of phage proteins during capsid assembly. Our investigation of regulatory evolution contributes to interpreting genome-level assessments of regulatory variation, as well as to understanding the molecular basis of adaptation.

## Background

Surveys of variation in gene regulation among different species, populations and individuals indicate that such variation may be strongly influenced by genetic drift and purifying selection [1-4]. This evolutionary pattern is reminiscent of the generally neutral evolution of synonymous sites and of purifying selection on non-synonymous sites in protein-coding genes [5]. Similar to positive selection observed at non-synonymous sites in protein-coding genes, changes in gene expression also have been shown to be subject to positive selection [6-8].

Investigating genes whose expression is affected by positive selection will allow us to understand the role of gene regulation in evolution [9-12].

The distinction between the molecular components that influence gene regulation that might evolve in a neutral fashion and those that may respond to positive or purifying selection is not clear [13-15]. Indeed, only a few studies have shown the adaptive value of point mutations in *cis*-acting regulatory sequences or *trans*-acting regulatory proteins let alone the effect of such adaptive substitutions upon gene regulation [16-19]. Additionally, genetic variation within promoter regions may affect chromatin structure leading to differences in gene expression [18]. Furthermore, *cis*-regulatory sites and *trans*-regulatory proteins may occur in a variety of

\* Correspondence: celesteb@uidaho.edu

<sup>1</sup>Department of Biological Sciences, PO Box 443051, University of Idaho, Moscow, ID 83843-3051 USA

Full list of author information is available at the end of the article

combinations with the same effect on gene expression [20]. This complexity makes it difficult to determine the genetic basis of regulatory variation or the forces that influence their evolution.

A reductionist approach in which adaptive substitutions in regulatory regions are tested for their influence on gene regulation and fitness provides another pathway for evaluating the role of gene regulation in evolution [4]. First, the adaptive substitution and its role in gene regulation as either a *cis*- or a *trans*-regulatory mutation are known. Second, the evolutionary forces that led to the substitution's detection are clearly defined by the experimental conditions under which it was discovered. Third, the effects of the substitution on gene regulation and on fitness are easily determined. In this approach, simple hypotheses can be easily defined and tested. For example, mutations that alter transcription factor binding sites such that binding efficiency is reduced are hypothesized to lead to the down regulation of gene transcription, but the environments in which such down regulation will improve fitness are unknown. These points can be addressed directly by a reductionist approach, providing a bottom up view of the evolutionary forces and genetic basis of regulatory variation.

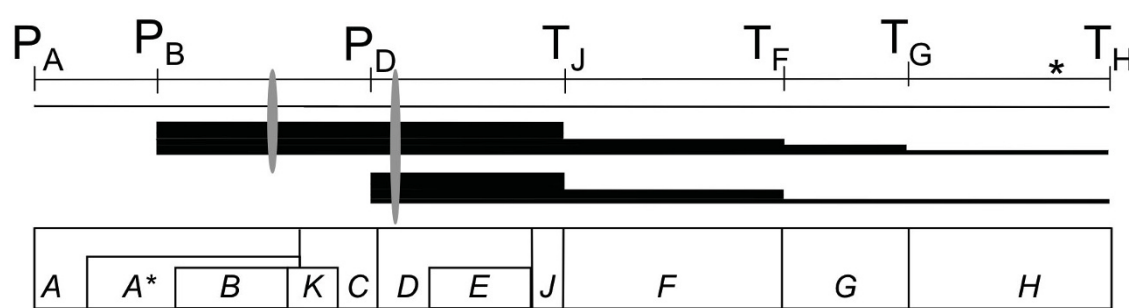
Such an approach requires a model organism whose biology and evolution are well studied. For over 50 years, the bacteriophage  $\phi$ X174 has been a model organism in the study of DNA replication, gene regulation and virion structure, so that the molecular biology and physiology of this positive-strand, ssDNA phage is well understood [21,22].  $\phi$ X174 utilizes its host's replication, transcription and translation machinery in order to reproduce making it an amenable organism for studies of regulatory evolution. In particular, the only components of the gene expression machinery that are encoded by the phage genome are the *cis*-acting regulatory sequences, although a possible role for the phage protein H was recently proposed [23]. Hence, studies of regulatory evolution using  $\phi$ X174 can focus on the importance of *cis*-acting sequences.

The simplicity of  $\phi$ X174 gene regulation also makes it a useful model for evolutionary studies of gene transcription. Over the approximately 12 minute life cycle of the wild type phage, 11 genes are expressed (Figure 1, bottom). These genes encode six procapsid proteins (B, D, F, G, H, J), two proteins that are involved in viral genome replication (A, C), two that interact with other phage or the host (A\*, E), and one whose function is unknown (K) [22,24]. Gene transcription is controlled by promoters that appear to consist solely of sigma factor binding sites upstream of genes A, B and D (Figure 1, top; P<sub>A</sub>, P<sub>B</sub>, P<sub>D</sub>) [25-27]. The transcript that starts from P<sub>A</sub> encodes the A and A\* proteins and is very unstable. The length of transcripts starting from P<sub>B</sub> and P<sub>D</sub> varies

based upon which of the rho-independent termination signals after genes J, F, G and H are used (Figure 1, top; T<sub>J</sub>, T<sub>F</sub>, T<sub>G</sub>, T<sub>H</sub>) [28,29]. Thus the transcripts that begin from P<sub>B</sub> and P<sub>D</sub> overlap in encoding six of nine genes (Figure 1, middle). The percentage of transcripts that continue through a termination signal to the following gene differs for each terminator (Figure 1, middle), and there is considerable variation in the half-life of each transcript [28,30]. A ribosome binding site immediately upstream of the start codon has been identified experimentally for genes A, D, E, F, G and H, and there are consensus Shine-Dalgarno sequences appropriately spaced upstream of B and J, but not before C [27,30,31].  $\phi$ X174 is an excellent model organism given this detail of understanding of its life cycle and the relative simplicity of its regulatory machinery.

For over 15 years, the bacteriophage  $\phi$ X174 has been a model organism for the study of the genetic basis of adaptation as well [32]. Many of these studies have grown phage in chemostats under various environmental conditions to observe evolution under strong positive selection with negligible influence of genetic drift [33,34]. Selection pressures in the chemostats include competition between phage and adaptation to novel temperatures and/or hosts, which are not allowed to evolve during these experiments. Multiple parallel substitutions arise under similar environmental conditions, especially in the major capsid protein F, the replication initiation protein A and the pilot protein H [32]. These chemostat experiments also have resulted in adaptive substitutions in the promoters before genes A, B and D, the ribosome binding sites of A, E and G, and the transcript termination loops after genes J and H [HA Wichman, personal communication; [35-38]]. Indeed, the fraction of sites undergoing adaptive substitutions in both regulatory sequences and protein sequences is the same, 11-12%, even though the target size for regulatory substitutions is an order of magnitude smaller than the total protein sequence of this phage [32]. Thus the adaptive substitutions identified in these experiments provide multiple opportunities for understanding the molecular and phenotypic consequences of regulatory evolution.

Our goal is to understand the processes involved in regulatory evolution by identifying adaptive substitutions in regulatory sequences, observing the influence these substitutions have upon gene regulation and correlating this regulatory variation to fitness. Our choice of regulatory substitutions is guided by their spread under experimental conditions that elicit strong positive selection. Therefore, we are focusing on adaptive substitutions that have accumulated in *cis*-regulatory sequences of the bacteriophage  $\phi$ X174 in response to selection at high temperatures in previous evolution



**Figure 1 Regulation of ϕX174 gene transcription.** Gene transcription in ϕX174 is regulated by four promoters (P<sub>A</sub>, P<sub>B</sub>, P<sub>D</sub>, \*) and four transcription terminators (T<sub>J</sub>, T<sub>F</sub>, T<sub>G</sub>, T<sub>H</sub>). Bars indicate transcript start and stop and thickness of bars indicates relative amount of transcript. Ovals show approximate positions of sequences detected in qPCR experiments. Gene locations are shown below, and genes B, K and E are in different reading frames from the genes with which they overlap. The ϕX174 genome is a circular molecule; the figure was linearized for clarity. From [21,22].

experiments [32]. In the present study, we find that these adaptive substitutions down-regulate gene expression at both temperatures and that they consistently increase fitness at high temperature, regardless of bacterial host.

Results

The goal of this study was to determine whether strong positive selection influences gene transcription. In multiple chemostat evolution experiments, adaptive substitutions arose in at least one of the three ϕX174 promoters. We chose to study the -35 region of the sigma factor binding site upstream of gene D (Figure 1; P<sub>D</sub>), because substitutions arose in this region in response to selection at high temperature in three hosts, *Escherichia coli*, *Shigella sonnei* and *Salmonella typhimurium* (Table 1) [HA Wichman, personal communication; [34-38]]. These substitutions arose within 10 days and as early as five days when grown on *S. typhimurium* at high temperature and within 20 days when grown on *E. coli* or *S. sonnei*. In only two experiments were more than one of these four substitutions detected in the population at the same time. Interestingly, in both cases

<sup>G</sup>319<sup>T</sup> was fixed in the population and <sup>A</sup>323<sup>G</sup> eventually was lost [34,37].

The gene for the DNA packaging protein C is immediately upstream of gene D and contains the entire D promoter sequence. Two of the substitutions, <sup>G</sup>319<sup>T</sup> and <sup>A</sup>323<sup>G</sup>, affect the amino acid sequence of protein C (<sup>V</sup>63<sup>F</sup>, <sup>N</sup>64<sup>G</sup>, respectively). Two of the substitutions, <sup>T</sup>321<sup>C</sup> and <sup>C</sup>324<sup>T</sup>, are silent substitutions. Additionally, <sup>G</sup>319<sup>T</sup> is immediately adjacent to the canonical -35 sigma factor binding site, rather than within the site, so the effect of nearby substitutions can be observed. Using site-directed mutagenesis of the ancestral ϕX174 sequence, four strains were constructed that had a single substitution at one of these four sites [35]. Quantitative PCR (qPCR) was used to measure the transcript levels from the D promoter for the ancestral and mutant strains under each environmental condition. Two fitness assays, time to lysis and doublings per hour, were conducted for each of the four mutant strains and for the ancestral strain at 37°C and 42°C in each of the three bacterial hosts. Chemostats were used to directly compete the ancestor with either of two mutant strains at 37°C or 42°C using *E. coli* C as the host.

**Table 1 Characteristics of adaptive substitutions in the D promoter of ϕX174 that arose in previous evolution experiments**

Strain	Mutation	Change in Protein C	# of Exps	Host*	Temperature, °C*
mut319	<sup>G</sup> 319 <sup>T</sup>	<sup>V</sup> 63 <sup>F</sup>	4**	<i>Salmonella typhimurium</i>	43.5
			2, 1	<i>Escherichia coli</i>	37, 43.5
mut321	<sup>T</sup> 321 <sup>C</sup>	none	1	<i>Salmonella typhimurium</i>	43.5
mut323	<sup>A</sup> 323 <sup>G</sup>	<sup>N</sup> 64 <sup>G</sup>	3	<i>Salmonella typhimurium</i>	43.5
			1	<i>Escherichia coli</i>	43.5
mut324	<sup>C</sup> 324 <sup>T</sup>	none	1	<i>Shigella sonnei</i>	42
			1	<i>Escherichia coli</i>	43.5

\*Conditions under which substitutions arose [HA Wichman, pers. comm.; [35-38]].  
\*\* Also, one reversion after switch to *E. coli* host at 43.5°C.

### Transcription is strongly down-regulated for all four mutants regardless of temperature or host

The promoter upstream of *B* controls the start of transcription for up to nine of eleven  $\phi$ X174 genes, including *C*, and these transcripts overlap with the transcripts for six genes that start at the promoter upstream of gene *D* (Figure 1). Amplification of cDNA for gene *D*, therefore, includes transcripts from both the *B* promoter and the *D* promoter. Thus, the transcript level of the *B* gene can serve as an internal control for the efficiency of infection, for contaminating phage DNA and for the effect of temperature or host on phage transcription in general. Additionally, the transcript levels for gene *B* are a proxy for the transcript levels for gene *C*, because transcripts other than those shown in Figure 1 are rapidly degraded and do not contribute to phage mRNA levels [28].

The effect of mutations in the promoter upstream of gene *D* was determined at two time points after the synchronized injection of phage DNA into host cells. After RNA extraction and conversion to cDNA, regions of the *B* and *D* cDNAs (Figure 1, grey ovals) were amplified by qPCR, and the ratio of the *D* transcript relative to the *B* transcript was measured (Figure 2). In each host at both temperatures and at both time points, the consequences of the mutations in the promoter were the same. All of the mutant strains were significantly down-regulated compared with the ancestor (Table 2). The decrease in the amount of transcript for the *D* gene ranged from no transcript to almost the same amount of transcript as for the *B* gene.

### Fitness of all mutant strains was higher than ancestor at high temperature

Fitness was measured as the number of times the phage population size doubles in an hour when grown in batch culture. Fitness was measured for each strain with each host at each temperature, and the fitnesses of the mutant strains were compared to that of the ancestor. In all three hosts, the mutants had more doublings per hour than the ancestor when grown at 42°C regardless of the host, but the amount of improvement of the mutant over the ancestor at 42°C depended upon the host (Figure 3, Table 3). At 37°C, relative fitnesses of mutants to ancestor depended upon the host. In *S. typhimurium*, all of the mutants had higher fitness than the ancestor ( $p = 0.001$ ). In *S. sonnei*, all of the mutants had lower fitness ( $p < 0.0001$ ); in *E. coli*, the mutants had the same or lower fitness than the ancestor at 37°C (Table 3).

### Mutants out-compete the ancestor in a chemostat at high temperature

The substitutions in the *D* promoter evolved in the high multiplicity of infection (MOI) environment of a

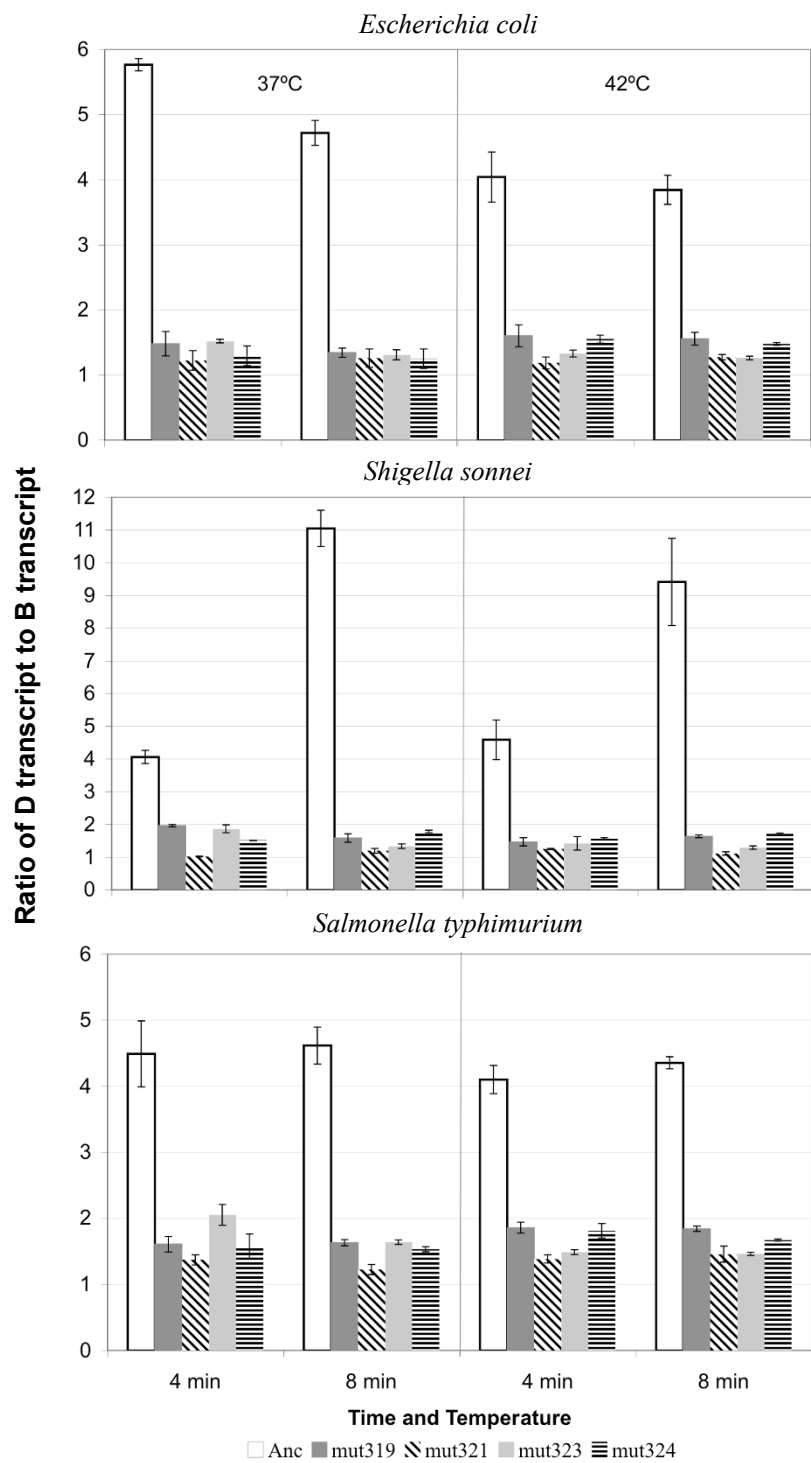
chemostat, but the fitness assays were conducted in low MOI batch cultures. Competition assays between the ancestor and either of two mutant strains were conducted in chemostats to confirm that the differences in fitness shown in Figure 3 reflect a competitive advantage within a chemostat. Figure 4 shows the average change in frequency of the mutant for replicate competition experiments in *E. coli* at two temperatures. At 37°C, the frequency of the mutant barely changes, even when the chemostat is started at different ratios of mutant to ancestor. At 42°C, however, the mutant frequency increases to around 80% over a two-hour experiment (Figure 4). In contrast with the low MOI fitness measurements, which show that doublings per hour of the mutant is less than the ancestor, the competition assays at 37°C indicate that the ancestor does not out-compete the mutants. The results of the competitions at 42°C, however, coincide with the low MOI fitness measurements. The change in frequency of the mutant strains in the 42°C chemostats is significantly greater than in the 37°C chemostats (Wilcoxon rank sum test,  $p < 0.005$ ).

### Time to lysis for all mutant strains is different from ancestor, except in *S. typhimurium*

One component of overall fitness is the time it takes from injection of the phage DNA into the cell until the cell is lysed, releasing the progeny phage. Phage that lyse earlier can infect new hosts earlier, but late-lysing phage may have more progeny. Time to lysis was measured for each of the five phage strains under six conditions: three hosts at two temperatures. The time point at which the number of phage in the culture was greater than or equal to twice the number at the zero time point determined time to lysis. This is a coarse-scale assessment because time points were taken every two minutes starting at ten minutes after eclipse of the phage into the host cell. This coarse scale was more than sufficient for seeing differences between phage in *E. coli* and *S. sonnei*. In both hosts, the ancestor lysed by 12 minutes at 37°C and by 14 minutes at 42°C. Each of the mutants, on the other hand, lysed by 14 minutes at 37°C and by 12 minutes at 42°C, reversing the pattern of the ancestor. This pattern was not seen in *S. typhimurium*, where only mut323 at 42°C showed an earlier lysis time of 14 minutes. All of the other phage lysed *S. typhimurium* by 16 minutes at both temperatures.

### Discussion

We have shown that substitutions in regulatory sequences contribute to adaptation under strong selective conditions. These adaptive substitutions down-regulate the expression of transcripts from the *D* promoter. Compared to the ancestor, the mutants had significantly lower transcript levels regardless of the host or



**Figure 2 Transcript levels differ between ancestor and promoter mutant strains at two temperatures in three bacterial hosts.** The amount of transcript starting from the D promoter was measured relative to the amount of transcript starting from the B promoter for each phage strain using quantitative PCR. Injection of DNA into host was synchronized by attachment at 15°C and then addition to LB at 37°C. Samples were taken at 4 and 8 min after injection, which is prior to the start of cell lysis. The y-axis indicates the ratio of the D transcript to the B transcript. Bars are the mean and error bars the standard error of two (*S. sonnei*) or three replicate assays. Solid bars are mutant strains with non-synonymous substitutions in C, hatched bars are synonymous in C. Relative transcript level of the ancestor is significantly greater than the four mutants for all host/temperature/time combinations (Table 2). Note the difference in scale for the *S. sonnei* graph.

**Table 2 Transcript level from the D promoter is significantly reduced in promoter mutant strains at any temperature, host or time point\*.**

Temperature	Host	Time	F-test	Pr > F
37°	<i>Escherichia coli</i>	4 min	395	<0.0001
37°	<i>Escherichia coli</i>	8 min	397	<0.0001
37°	<i>Salmonella typhimurium</i>	4 min	53	<0.0001
37°	<i>Salmonella typhimurium</i>	8 min	209	<0.0001
37°	<i>Shigella sonnei</i>	4 min	450	<0.0001
37°	<i>Shigella sonnei</i>	8 min	1112	<0.0001
42°	<i>Escherichia coli</i>	4 min	143	<0.0001
42°	<i>Escherichia coli</i>	8 min	390	<0.0001
42°	<i>Salmonella typhimurium</i>	4 min	381	<0.0001
42°	<i>Salmonella typhimurium</i>	8 min	473	<0.0001
42°	<i>Shigella sonnei</i>	4 min	94	0.0002
42°	<i>Shigella sonnei</i>	8 min	143	<0.0001

\*ANOVA on relative transcript levels for each environmental condition and time point with 1 degree of freedom for each test.

temperature at which the phage were propagated. This change in expression, however, was only consistently correlated with adaptation in the high temperature conditions under which these substitutions were first identified.

One reason for choosing the -35 sequence of the *D* promoter was because of the number of sites at which adaptive substitutions arose in this region and because two of the four substitutions were synonymous, <sup>T321C</sup> and <sup>C324T</sup>, and two non-synonymous, <sup>G319T</sup> and <sup>A323G</sup>, in protein C (Table 1). Thus, we are able to observe the fitness effect of the changes in transcription alone and the changes in transcription and in the amino acid sequence of C. The two synonymous substitutions appear to have almost identical doublings per hour for each experimental condition (Figure 3). The two non-synonymous substitutions, on the other hand, do not appear to be similar to each other or to the synonymous substitutions. There appear to be pleiotropic effects on fitness due to the amino acid substitutions in protein C for the two non-synonymous substitutions, as we expected.

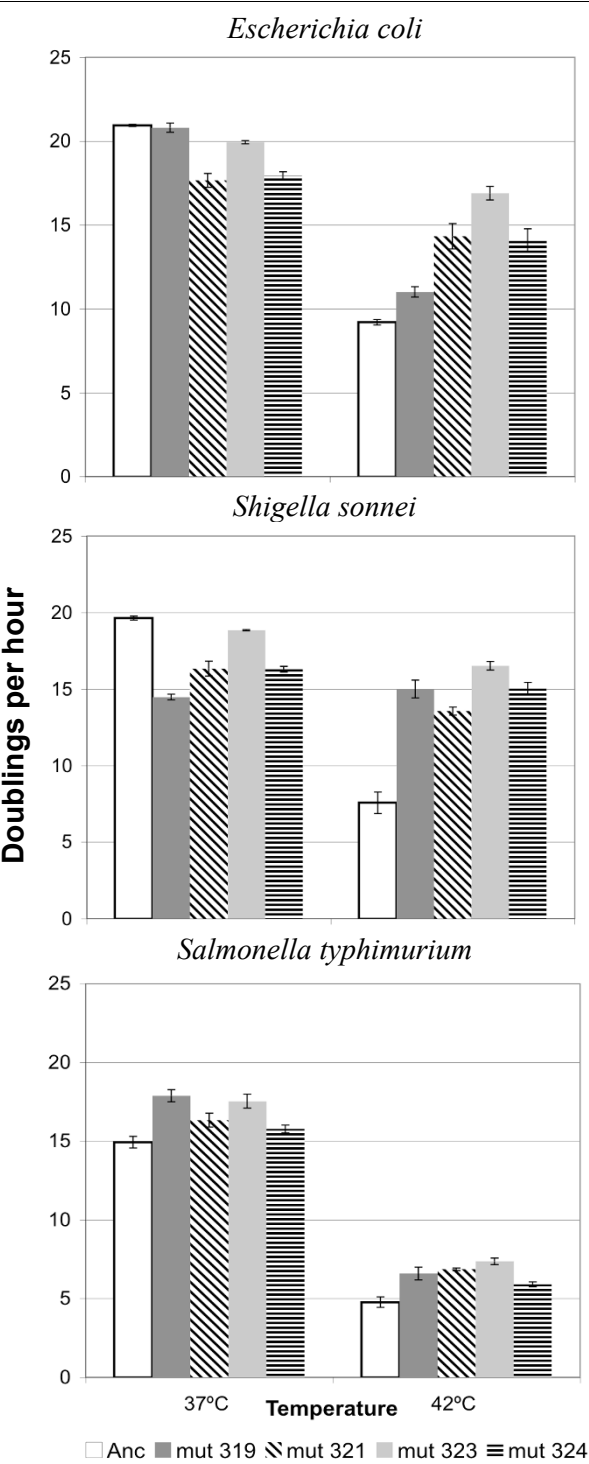
Although synonymous substitutions are known to be adaptive under certain circumstances, these circumstances are not applicable here. The synonymous substitutions, <sup>T321C</sup> and <sup>C324T</sup>, are for common codons in the *E. coli* C genome. Moreover, the correlation between codon bias and selective advantage is generally correlated with highly-expressed genes, and protein C has a very low level of expression [39]. Indeed, there is no canonical Shine-Dalgarno sequence before C's translation start codon. It is possible that all four substitutions affect some other aspect of phage biology, such as genome packaging or DNA secondary structure, however, the inconsistency in the effect on GC-content suggests

that these are also unlikely factors. The strict decrease in transcript levels for all conditions and the significant increase in fitness at high temperature for all hosts suggests to us that transcription provides the underlying basis for the adaptive advantage of these synonymous substitutions.

It was somewhat surprising to us that each of the four substitutions would have such a strong effect on fitness on their own because we expected epistatic effects to be important.  $\phi$ X174 was lab-adapted to *E. coli* starting in the 1950's, and its fitness on *S. typhimurium* is quite low (Figure 3). Multiple experiments have shown that the first adaptive substitutions to sweep through a chemostat in which  $\phi$ X174 is grown on *S. typhimurium* at 43.5°C generally are in the capsid protein F, and these substitutions seem to affect capsid stability and/or host attachment [38,40]. These selective sweeps are accompanied by a large increase in the fitness of the evolved phage. Either the <sup>G319T</sup> or the <sup>A323G</sup> substitutions were subsequently seen in these same populations, and <sup>G319T</sup> and several of the adaptive substitutions in the capsid protein reverted to the ancestral state when the strain carrying them was grown on *E. coli* [36,37]. These results suggest that the amino acid substitutions in protein C (<sup>V63F</sup>, <sup>N64G</sup>) are interacting either with the host or with the capsid protein. The small increase in fitness for our single mutants when *S. typhimurium* was the host (Figure 3) further suggests that the interaction is with the *Salmonella*-adapted capsid protein. We hypothesize that had adaptive substitutions been present in the capsid proteins of the mutant strains, the increase in fitness due to the regulatory substitutions would have been more substantial. In particular, because <sup>G319T</sup> has occurred in the most high temperature, *Salmonella* experiments, we expect that it has the greatest fitness effect in a *Salmonella*-adapted background contrary to its small fitness advantage relative to the other mutants. Tests for such epistatic and pleiotropic interactions await future experiments.

We can think of several non-exclusive hypotheses to test our conclusion that down regulating gene expression is adaptive at high temperature. First, the higher temperature is deleterious for the host as well as the phage. The main costs of constitutive expression of non-essential genes in *E. coli* are the acts of transcription and/or translation, and over expression of genes is also costly in yeast [41,42]. Since the phage use the transcription and translation machinery of the host, down regulating gene expression may reduce the stress on the host's regulatory machinery, thus maximizing the number of phage progeny produced.

Secondly, it seems reasonable to predict that the ratio of the different transcripts (Figure 1) reflects the levels needed to produce the optimal ratio of proteins for each



**Figure 3 Fitness differences between ancestor and promoter mutant strains at two temperatures for three bacterial hosts.** Doublings per hour were measured for each phage strain in three bacterial hosts and at two temperatures. Bars are the mean and error bars the standard error of three replicate assays. Solid bars are mutant strains with non-synonymous substitutions in C, hatched bars are synonymous in C. Doublings per hour of the ancestor is significantly lower than at least one of the four mutant strains for all host/temperature combinations except *E. coli* and *S. sonnei* at 37°C, where the ancestor's fitness is significantly greater than at least one of the four mutant strains (Table 3).

**Table 3 Fitness of promoter mutants is greater than ancestor at 42°C\*.**

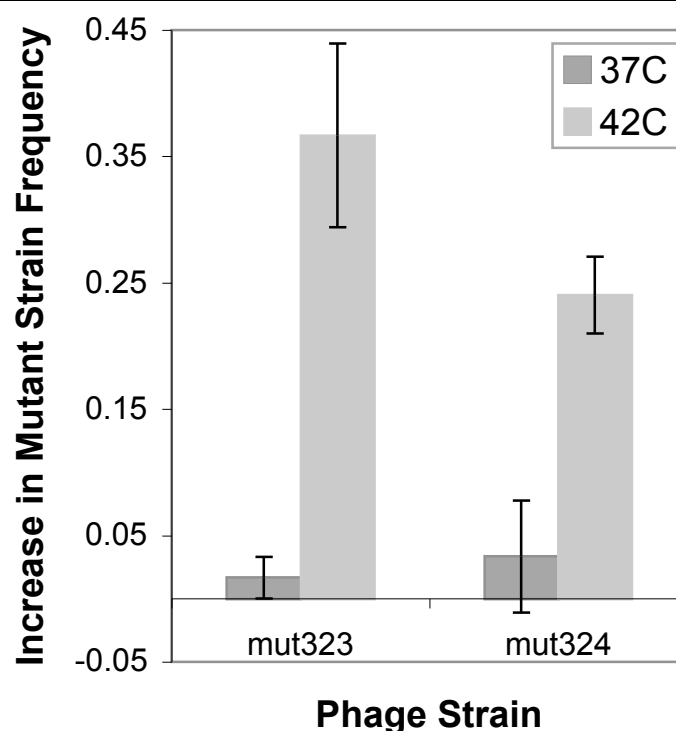
Temperature	Host	F-test	Pr > F	t-value	Pr >  t
37°	<i>Escherichia coli</i>	41.4	<0.0001	6.7	<0.0001
37°	<i>Salmonella typhimurium</i>	10.6	0.0013	-4.6	0.0010
37°	<i>Shigella sonnei</i>	70.6	<0.0001	11.3	<0.0001
42°	<i>Escherichia coli</i>	35.4	<0.0001	-8.6	<0.0001
42°	<i>Salmonella typhimurium</i>	14.8	0.0003	-6.6	<0.0001
42°	<i>Shigella sonnei</i>	53.2	<0.0001	-13.9	<0.0001

\*ANOVA on doublings per hour for each environmental condition with 4 degrees of freedom for each test, and t-test comparing ancestor to all mutants.

capsid. The  $\phi$ X174 procapsid is composed of 60 B (internal scaffold), 60 F (major capsid), 60 G (major spike), 12 H (pilot) and 240 D (external scaffold) proteins. In the provirion, the B proteins are replaced by 60 J (DNA binding) proteins [21]. The reduction in transcript levels due to the three transcription terminators prior to gene *H*, which has the lowest number of subunits per procapsid, and the strong promoter before gene

*D*, which has the highest number of subunits per procapsid, lend support for this prediction (Figure 1). Environmental changes that disrupt the ratio of phage proteins, such as increased temperature, may be a selective force leading to changes in gene regulation.

Our second hypothesis, therefore, is that the relative concentrations of proteins at high temperature are sub-optimal, possibly because some of the proteins are unstable or because the efficiency of the transcription terminators changes at the higher temperature. Protein B is required for the assembly of pentamers of protein F, for the assembly of pentamers of protein G and pentamers of protein F into an intermediary complex, and B facilitates the incorporation of protein H into this complex [43,44]. Protein B appears to have a chaperone-like quality in that it keeps protein F as well as protein H from aggregating [23,45]. If the amount of protein B that is produced at 42°C is insufficient for performing these functions, then decreasing the amount of proteins F and H by down regulating transcription from the *D* promoter may optimize the ratio of phage proteins at this temperature. Our future work will test whether maintaining the proper ratio of phage proteins is



**Figure 4 Competition assays confirm the greater fitness of mutant strains at 42°C in *E. coli*.** The ancestor and one mutant strain were grown together in a two-stage chemostat on *E. coli* C at 37°C or 42°C. Concentration of each strain was determined by allele specific PCR at the beginning and end of each chemostat. The y-axis shows the average increase in frequency of the mutant strain after two hours. Three replicate chemostats were run for each mutant at each temperature. Mut323 is a non-synonymous substitution in *C*, and mut 324 is synonymous. The increase in mutant strain frequency was significantly different between the two temperatures for both mutants (Wilcoxon rank sum test,  $p < 0.005$ ).



the molecular basis of these adaptive regulatory substitutions.

These two hypotheses do not shed light on the puzzling inconsistency between the time to lysis studies and the lower transcription rates of the mutants. The lysis protein E is a transmembrane protein that affects the host so that at cell division, the cell membrane fails and lysis occurs [46]. It seems consistent that at 37°C, the ancestor would lyse earlier than the mutants, which are producing less transcript from the *D* promoter, and hence, fewer transcripts for protein E (Figure 1). It is not consistent that at 42°C when the transcription level of the mutants is again low relative to the ancestor, that the time to lysis should decrease for the mutants. Our first hypothesis seems to be the most consistent with the results at 42°C. If the host's transcription/translation machinery are overtaxed at high temperature, the length of the cell cycle may be increased. Decreasing the pressure on this machinery by down regulating gene expression may lead to earlier cell division, and thus earlier lysis.

## Conclusions

One consequence of strong positive selection on the bacteriophage  $\phi$ X174 was to down regulate gene transcription from one of two promoters whose transcripts overlap in encoding six of nine genes. The fitness advantage that correlated with this change in transcript levels was consistent only under high temperature conditions, suggesting that understanding the evolution of gene regulation will rely upon evaluating regulatory variants under the appropriate environmental conditions. The physiological mechanism by which decreased transcript levels may lead to increased fitness is still to be ascertained. Continued in-depth studies on the molecular basis of adaptation by changes in gene regulation will provide essential background for following up on whole-genome expression studies.

## Methods

### Phage and Hosts

A wild-type isolate of  $\phi$ X174 was used as the ancestor for these studies (GenBank Accession AF176034). The hosts used for phage propagation were *Escherichia coli* C, *Salmonella typhimurium* LT2 strain IJ750 [*xyl*-404 *met*A22 *met*E551 *gal*E719 *trp*D2 *ilv*-452 *hsd*LT6 *hsd*SA29 *hsd*SB1 *fla*-66 *rps*L120 *H1*-b *H2*-c *nix*], and *Shigella sonnei* strain NCDC 1120-66 [CIP 104223]; (ATCC #25931).

### Site-directed Mutagenesis

Four mutant strains were constructed in the ancestral background using site-directed mutagenesis. Each half of the phage was amplified using Platinum Pfx DNA

Polymerase from Invitrogen (Cat. No. 11708) following the manufacturer's directions. The forward mutant primers were paired with 2953R and the reverse mutant primers were paired with 2605F in separate reactions (Table 4). The two products for each mutant were combined in another PCR to join the two halves, and this reaction was electroporated into competent *E. coli* C cells. The electroporated cells were plated and plaques were purified by a second round of plating. The entire genomes of each purified plaque were sequenced to confirm the presence of the desired mutation and no others. Each mutant strain had a single mutation in the *D* promoter: mut319 (<sup>G</sup>319<sup>T</sup>), mut321 (<sup>T</sup>321<sup>C</sup>), mut323 (<sup>A</sup>323<sup>G</sup>), or mut324 (<sup>C</sup>324<sup>T</sup>).

### RNA Sampling and Purification

To synchronize injection of the phage into the host cells,  $1.5 \times 10^8$  cells were incubated with  $1.5 \times 10^7$  phage in a 1 ml eppendorf tube at 15°C (10°C for *Salmonella*) for one hour. The cells and phage were centrifuged in a benchtop micro-centrifuge at 13000 rpm for 5 minutes at 4°C. The supernatant containing unattached phage was discarded and pellets were resuspended in 1 ml of ice cold LB + 2 mM CaCl<sub>2</sub>. The resuspended pellet was added to 9 mls of LB + 2 mM CaCl<sub>2</sub> pre-equilibrated at either 37°C or 42°C; this procedure brings the culture above 22°C, at which point attached phage inject their DNA into the host. After the attached phage and cells were added to the temperature-equilibrated media, four and eight minute samples of 500  $\mu$ l were added directly to 1 ml of RNAprotect™ Bacteria Reagent (Qiagen Cat. No.76506) and incubated at room temperature for 5 minutes to stabilize the RNA. The samples were centrifuged for 10 minutes in a micro-centrifuge at room temperature, and total RNA was purified using the RNeasy Mini Kit (Qiagen Cat. No.74104). The optional on-column DNase digestion using the RNase-Free, DNase Set (Qiagen Cat. No. 79254) was performed on filter following the guidelines in the Qiagen protocol to remove contaminating DNA.

### qPCR Primers and Probes

Primers and 5' FAM™ dye-labeled Taqman® MGB probes with non-fluorescent quencher dyes were designed and optimized by Applied Biosystems (Custom Taqman® Gene Expression Assays Part No. 4331348) using their Custom Taqman® Assay Design Tool available on their website. The primer and probe sets detected cDNAs for genes *B*, nt 5075-49, or *D*, nt 390-848 (Table 4). They were tested for amplification efficiency using five 10-fold dilutions of the cDNA to produce a PCR standard curve, and the C<sub>T</sub> values were plotted against log input nucleic acid mass. A 100% efficient reaction will yield a 10-fold increase in PCR amplicon every 3.32 cycles

**Table 4 Oligonucleotides used for site-directed mutagenesis, allele-specific PCR and quantitative PCR**

Primer Name	Location*	Sequence (5' to 3')	Purpose
319F	314-338	CTCTTTTGACATTTTAAAGAGCG**	Mutagenesis
319R	338-314	CGCTCTTTTAAATGTCAAAGAG	Mutagenesis
321F	314-338	CTCTTGTCGACATTTTAAAGAGCG	Mutagenesis
321R	338-314	CGCTCTTTTAAATGTGCAAGAG	Mutagenesis
323F	314-338	CTCTTGTTGGCATTTTAAAGAGCG	Mutagenesis
323R	338-314	CGCTCTTTTAAATGCCAACAAGAG	Mutagenesis
324F	314-338	CTCTTGTTGATATTTTAAAGAGCG	Mutagenesis
324R	338-314	CGCTCTTTTAAATATCAACAAGAG	Mutagenesis
2605F	2605-2626	CAGGTTGTTCTGTTGGTGCTG	Amplification
2953R	2937-2953	CCGCCAGCAATAGCACC	Amplification
Anc_asF	305-324	GTAGAGATTCTCTTGTGAC	Allele-specific
mut323_asF	305-323	GTAGAGATTCTCTTGTGG	Allele-specific
mut324_asF	304-324	GGTAGAGATTCTCTTGTGAT	Allele-specific
phiXpD_asR	625-603	GCAATAAACTCAACAGGAGCAGG	Amplification
phiXpB_F	5337-5360	CTCAAATTTATGCGCGCTTCGATA	qPCR Primer
phiXpB_M	5377-7	CCAACCTGCAGAGTTT	qPCR Probe
phiXpB_R	27-7	TTCTGCGTCATGGAAGCGATA	qPCR Primer
phiXpD_F	400-424	TACTGAACAATCCGTACGTTTCCA	qPCR Primer
phiXpD_M	427-443	CCGCTTGGCCTCTATT	qPCR Probe
phiXpD_R	466-457	ACGGCAGAAGCCTGAATGAG	qPCR Primer

\* Based upon Sanger, et al. sequence [24].

\*\* Mutation underlined.

during the exponential phase (Applied Biosystems protocol). Two dilutions closest to 100% efficiency,  $10^{-1}$  and  $10^{-2}$ , were used in all subsequent experiments. Since *E. coli* genomic DNA and RNA are present in the samples, we tested the primers and probes designed for genes *B* and *D* of the bacteriophage for amplification of purified *E. coli* DNA. We also tested the cDNA samples purified with only bacteria in the sample for the same reason, and no amplification was found in either case.

#### Reverse Transcription and Real-time PCR

Reverse transcription was carried out with High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems #4368813) according to the manufacturer's protocol. The amplification and detection of the cDNA were carried out using Applied Biosystems 7900HT Fast Real-Time PCR Sequence Detection System in a 96-Well Optical Reaction Plate (Applied Biosystems). Thermal cycling conditions started with 95°C hold for 10 minutes followed by 40 cycles of 95°C for 15 seconds then 60°C for 1 minute. For determining the  $C_T$  values, the baseline was set to automatic and the threshold values were all manually set at 0.2. Positive controls used each set of primers against the same preparation of phage DNA. Negative controls included using each set of primers on the RNA to test for contaminating DNA and on a no target reaction mix. The ratio of *D* transcript relative to the *B* transcript was determined by

raising 2 to the difference between the  $C_T$  values for the *B* and *D* transcripts.

#### Fitness Assay

Growth rate was tested as an index for fitness for each strain [34]. Host cells were grown to a density of  $\sim 2 \times 10^8$  cells/mL by shaking at 37°C in 125 ml flasks with 10 mL LB supplemented with 2 mM  $\text{CaCl}_2$ . The cultures were moved to the experimental temperature, and phage were added at an initial concentration of  $\sim 10^4$ /mL. Samples were titrated at two time points, 0 min and 40 min. Growth rate in doublings per hour were calculated as:  $[\log_2(\text{phage concentration at } t = 40) - \log_2(\text{phage concentration at } t = 0)](60 \text{ min}/40 \text{ min})$ .

#### Chemostat Competition Assay

The ancestral strain was allowed to compete with either of two mutant strains, mut323 or mut324, in a two-stage chemostat. The chemostat consisted of 2  $100 \times 15$  mm glass test tubes. The first tube contained *E. coli* C with a continuous supply of LB broth containing 0.005% antifoam B (Astoria-Pacific Product # 90-0703-01) to maintain cell growth. The mixture in the cell tube was drawn into the second tube that contained the phage. Approximately 2 mls was maintained in each tube with a flow rate of approximately 9 ml/hr. The *E. coli* was grown in LB at 37°C for one hour with shaking prior to addition to the chemostat. Four mls of these

cells were added to the cell tube and allowed to flow into the phage tube. The amount was allowed to level out at 2 mls once again. One of the mutants and the ancestor were added to the chemostat at various concentrations ( $10^8$  total phage) with a syringe, and the chemostat was run at either 37°C or 42°C for two hours. The chemostat was run for two hours to minimize the possibility of mutations arising and sweeping through the chemostat. Samples were taken every fifteen minutes using a port in the phage tube, and the phage in the chemostat were titered at each of these time points to confirm that no such sweep occurred as indicated by a sudden rapid increase in titer. Allele-specific PCR (asPCR) was used to amplify 20 plaques from the mixture of phage added to the chemostat and 20 plaques from the final chemostat sample to determine the frequency of the ancestor and the mutant strains at the beginning and end of each competition.

#### Allele-specific PCR

For each plaque purified from the competition experiments, separate asPCR assays were performed for each phage strain included in the chemostat using primer pairs that specifically amplify only one strain (Table 4). The specificity of each allele-specific primer pair was confirmed for every PCR run by the inclusion of both positive and negative controls. Except for the optimized annealing temperatures, which were 63.1°C for mut323 and mut324 and 63.5°C for Ancestor, PCR was performed using Amplitaq Gold PCR Master Mix (Applied Biosystems) according to the manufacturer's protocol for a three-temperature cycle. PCR products were visualized on agarose gels stained with ethidium bromide and each reaction was scored for presence or absence of amplification.

#### Time to Lysis Assay

To find the time to lysis for the ancestor and each of the D promoter mutant strains, the samples were prepared as described for the transcription experiments except that  $10^7$  cells were combined with  $10^6$  phage. After attached phage were brought to the experimental temperature, samples were taken every two minutes for up to 24 minutes, plated on LB agar plates, and the plaques were counted after 3-4 hours of incubation at the experimental temperature. The time point at which the number of phage in the culture was greater than or equal to twice the number at the zero time point determined time to lysis.

#### Statistical Analyses

A three-factor ANOVA was conducted on the data for doublings per hour and a four-factor ANOVA was conducted for D transcript level to test whether the mutant strains differed significantly from the ancestral strain.

Because the other factors, host, temperature and (time), all showed significant interactions with strain, a one way ANOVA was conducted on each experimental condition separately. The null hypothesis for each combination of host, temperature and (time) was that the four mutant strains did not differ significantly from the ancestral strain.

#### Acknowledgements

Funding for this research was provided by NIH grant NCRR P20RR16448. We wish to thank Dr. Holly Wichman for providing the ancestral phage, bacterial host stocks and advice, Dr. Bentley Fane for advice and the University of Idaho phage group for their helpful assistance and input.

#### Author details

<sup>1</sup>Department of Biological Sciences, PO Box 443051, University of Idaho, Moscow, ID 83843-3051 USA. <sup>2</sup>Department of Biology, San Diego State University, MC4614, 5500 Campanile Drive, San Diego, CA 92182 USA.

#### Authors' contributions

CJB conceived of the study, participated in its design and coordination and drafted the manuscript. LZ conducted fitness and transcription assays in *Shigella*; KJE conducted fitness assays in *E. coli* and *Salmonella*; DA developed the allele-specific PCR assay; ADS conducted fitness, transcription and competition assays in *Salmonella* and/or *E. coli*. All authors read, provided intellectual input on and approved the final manuscript.

Received: 17 September 2010 Accepted: 3 December 2010

Published: 3 December 2010

#### References

- Jordan IK, Marino-Ramirez L, Koonin EV: Evolutionary significance of gene expression divergence. *Gene* 2005, **345**(1):119-126.
- Wang Y, Rekaya R: A comprehensive analysis of gene expression evolution between humans and mice. *Evol Bioinform Online* 2009, **5**:81-90.
- Tirosh I, Barkai N: Evolution of gene sequence and gene expression are not correlated in yeast. *Trends Genet* 2008, **24**(3):109-113.
- Gilad Y, Oshlack A, Rifkin SA: Natural selection on gene expression. *Trends Genet* 2006, **22**(8):456-461.
- Li WH, Wu CI, Luo CC: A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol* 1985, **2**(2):150-174.
- Emerson JJ, Hsieh LC, Sung HM, Wang TY, Huang CJ, Lu HH, Lu MY, Wu SH, Li WH: Natural selection on cis and trans regulation in yeasts. *Genome Res* 2010, **20**(6):826-836.
- Fraser HB, Moses AM, Schadt EE: Evidence for widespread adaptive evolution of gene expression in budding yeast. *Proc Natl Acad Sci USA* 2010, **107**(7):2977-2982.
- Whitehead A, Crawford DL: Variation within and among species in gene expression: raw material for evolution. *Mol Ecol* 2006, **15**(5):1197-1211.
- Cooper TF, Rozen DE, Lenski RE: Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc Natl Acad Sci USA* 2003, **100**(3):1072-1077.
- Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, DeSevo CG, Botstein D, Dunham MJ: The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet* 2008, **4**(12):e1000303.
- Fong SS, Joyce AR, Palsen BO: Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res* 2005, **15**(10):1365-1372.
- Wang L, Spira B, Zhou Z, Feng L, Maharjan RP, Li X, Li F, McKenzie C, Reeves PR, Ferenci T: Divergence involving global regulatory gene mutations in an *Escherichia coli* population evolving under phosphate limitation. *Genome Biol Evol* 2010, **2**:478-487.
- Lynch VJ, Wagner GP: Resurrecting the role of transcription factor change in developmental evolution. *Evolution* 2008, **62**(9):2131-2154.
- Wray GA: The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* 2007, **8**(3):206-216.

15. Emerson JJ, Li WH: **The genetic basis of evolutionary change in gene expression levels.** *Philos Trans R Soc Lond B Biol Sci* 2010, **365**(1552):2581-2590.
16. Chang YW, Robert Liu FG, Yu N, Sung HM, Yang P, Wang D, Huang CJ, Shih MC, Li WH: **Roles of cis- and trans-changes in the regulatory evolution of genes in the gluconeogenic pathway in yeast.** *Mol Biol Evol* 2008, **25**(9):1863-1875.
17. Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB: **Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila.** *Nature* 2005, **433**(7025):481-487.
18. Tirosch I, Weinberger A, Bezalel D, Kaganovich M, Barkai N: **On the relation between promoter divergence and gene expression evolution.** *Mol Syst Biol* 2008, **4**:159.
19. Wittkopp PJ, Haerum BK, Clark AG: **Evolutionary changes in cis and trans gene regulation.** *Nature* 2004, **430**(6995):85-88.
20. Wang QF, Prabhakar S, Chanan S, Cheng JF, Rubin EM, Boffelli D: **Detection of weakly conserved ancestral mammalian regulatory sequences by primate comparisons.** *Genome Biol* 2007, **8**(1):R1.
21. Fane BA, Brentlinger KL, Burch AD, Chen M, Hafenstein S, Moore E, Novak CR, Uchiyama A: **phiX174 et al. The Microviridae.** In *The Bacteriophages*. 2 edition. Edited by: Calendar R. New York: Oxford Press; 2006:129-145.
22. Hayashi M, Aoyama A, Richardson DL, Hayashi MN: **Biology of the Bacteriophage phiX174.** In *The Bacteriophages. Volume II*. Edited by: Calendar R. New York: Plenum Press; 1988:1-71.
23. Ruboyianes MV, Chen M, Dubrava MS, Cherwa JE Jr, Fane BA: **The expression of N-terminal deletion DNA pilot proteins inhibits the early stages of phiX174 replication.** *J Virol* 2009, **83**(19):9952-9956.
24. Sanger F, Coulson AR, Friedmann T, Air GM, Barrell BG, Brown NL, Fiddes JC, Hutchison CA, Slocumbe PM, Smith M: **The nucleotide sequence of bacteriophage phiX174.** *J Mol Biol* 1978, **125**(2):225-246.
25. Arquint M, Spencer JH: **Mapping the B and D gene promoters of bacteriophage S13 by footprinting and exonuclease III analysis.** *J Virol* 1987, **61**(12):3694-3700.
26. Arquint M, Spencer JH: **Mapping of RNA polymerase binding sites in the H/A gene region of bacteriophage S13 by footprinting and exonuclease III analysis.** *Biochim Biophys Acta* 1994, **1218**(3):339-345.
27. Sorensen SE, Barrett JM, Wong AK, Spencer JH: **Identification of the in vivo promoters of bacteriophages S13 and phi X174 and measurement of their relative activities.** *Biochem Cell Biol* 1998, **76**(4):625-636.
28. Hayashi M, Fujimura FK, Hayashi M: **Mapping of in vivo messenger RNAs for bacteriophage phiX-174.** *Proc Natl Acad Sci USA* 1976, **73**(10):3519-3523.
29. Hayashi MN, Hayashi M, Imai M: **Bacteriophage phi X174-specific mRNA synthesis in cells deficient in termination factor rho activity.** *J Virol* 1981, **38**(1):198-207.
30. Hayashi MN, Yaghamai R, McConnell M, Hayashi M: **mRNA stabilizing signals encoded in the genome of the bacteriophage phi x174.** *Mol Gen Genet* 1989, **216**(2-3):364-371.
31. Blasi U, Nam K, Lubitz W, Young R: **Translational efficiency of phi X174 lysis gene E is unaffected by upstream translation of the overlapping gene D reading frame.** *J Bacteriol* 1990, **172**(10):5617-5623.
32. Wichman HA, Brown CJ: **Experimental evolution of viruses: Microviridae as a model system.** *Philos Trans R Soc Lond B Biol Sci* 2010, **365**:2495-2501.
33. Bull J, Millstein J, Orcut J, Wichman H: **Evolutionary feedback mediated through population density, illustrated with viruses in chemostats.** *American Naturalist* 2006, **167**:E39-E51.
34. Wichman HA, Millstein J, Bull JJ: **Adaptive molecular evolution for 13,000 phage generations: a possible arms race.** *Genetics* 2005, **170**(1):19-31.
35. Bull JJ, Badgett MR, Wichman HA, Huelsenbeck JP, Hillis DM, Gulati A, Ho C, Molineux U: **Exceptional convergent evolution in a virus.** *Genetics* 1997, **147**(4):1497-1507.
36. Crill WD, Wichman HA, Bull JJ: **Evolutionary reversals during viral adaptation to alternating hosts.** *Genetics* 2000, **154**(1):27-37.
37. Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ: **Different trajectories of parallel evolution during viral adaptation.** *Science* 1999, **285**(5426):422-424.
38. Wichman HA, Scott LA, Yarbey CD, Bull JJ: **Experimental evolution recapitulates natural evolution.** *Philos Trans R Soc Lond B Biol Sci* 2000, **355**(1403):1677-1684.
39. Sharp PM, Emery LR, Zeng K: **Forces that influence the evolution of codon bias.** *Philos Trans R Soc Lond B Biol Sci* 2005, **360**(1544):1203-1212.
40. Bull JJ, Badgett MR, Wichman HA: **Big-benefit mutations in a bacteriophage inhibited with heat.** *Mol Biol Evol* 2000, **17**(6):942-950.
41. Stoebe DM, Dean AM, Dykhuizen DE: **The cost of expression of Escherichia coli lac operon proteins is in the process, not in the products.** *Genetics* 2008, **178**(3):1653-1660.
42. Wagner A: **Energy costs constrain the evolution of gene expression.** *J Exp Zool B Mol Dev Evol* 2007, **308**(3):322-324.
43. Burch AD, Fane BA: **Genetic analyses of putative conformation switching and cross-species inhibitory domains in Microviridae external scaffolding proteins.** *Virology* 2003, **310**(1):64-71.
44. Novak CR, Fane BA: **The functions of the N terminus of the phiX174 internal scaffolding protein, a protein encoded in an overlapping reading frame in a two scaffolding protein system.** *J Mol Biol* 2004, **335**(1):383-390.
45. Siden EJ, Hayashi M: **Role of the gene beta-product in bacteriophage phi-X174 development.** *J Mol Biol* 1974, **89**(1):1-16.
46. Bernhardt TG, Roof WD, Young R: **Genetic evidence that the bacteriophage phi X174 lysis protein inhibits cell wall synthesis.** *Proc Natl Acad Sci USA* 2000, **97**(8):4297-4302.

doi:10.1186/1471-2148-10-378

**Cite this article as:** Brown et al.: Positive selection at high temperature reduces gene transcription in the bacteriophage phiX174. *BMC Evolutionary Biology* 2010 **10**:378.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

